第40卷 第11期
2023年11月

干 旱 区 研 究
ARID ZONE RESEARCH

Vol. 40 No. 11
Nov. 2023

基于多植被指数组合的棉花叶片叶绿素含量估算

阿热孜古力·肉孜[1,2,3]， 买买提·沙吾提[1,2,3]， 何旭刚[1,2,3]， 冶晓文[1,2,3]

（1. 新疆大学地理与遥感科学学院，新疆 乌鲁木齐 830017；2. 新疆绿洲生态重点实验室，新疆 乌鲁木齐 830017；3. 智慧城市与环境建模自治区普通高校重点实验室，新疆 乌鲁木齐 830017）

摘 要：叶绿素含量是表征植被生长状况的重要参考指标，利用高光谱技术快速，精确地监测棉花叶片叶绿素含量，以新疆125个苗期棉花叶片样本为研究对象，通过测定其叶绿素含量与光谱数据，采用多种光谱预处理和多植被指数相结合的方法，构建了WOA-RFR棉花叶片叶绿素含量定量反演模型，并与SVR和RFR模型结果进行对比分析。结果表明：(1) 光谱变换方法中对数变换、分数阶微分和连续小波变换均能有效地提高植被指数与叶绿素含量的相关性。(2) 基于分数阶微分0.9阶变换的Vogelmann3、RVI、DVI、$SR_{[675-700]}$、$Mndvi_{705}$、ND、VOG1、NVI、TVI和VOG2植被指数组合的WOA-RFR模型反演效果最佳，其建模集和验证集模型R^2分别为0.920和0.955，RMSE分别为0.987和0.986，MRE分别为0.013和0.014，与RFR和SVR模型相比，预测精度有所提高，WOA算法优化模型效果明显。研究结果可为棉花叶片叶绿素含量定量反演提供决策依据。

关键词：植被指数组合；棉花；叶绿素含量；鲸鱼优化算法

叶绿素含量是表征作物健康状况的重要指标，其与作物生长紧密相关[1]。叶绿素含量的传统测定方法成本高、操作复杂、效率低，难以满足精准农业快速发展的需求[2]。目前，精确、快速、高效的高光谱技术被证明在植被叶绿素含量获取方面有一定的优势[2]。当前棉花叶片叶绿素含量的光谱反演主要采用多种预处理方式相结合的方法，分析光谱与叶绿素含量的相关性，并选择特征波长或者构建高光谱植被指数，建立叶绿素含量监测模型[1,3-4]。前人利用多种光谱预处理方法，如小波变换[5]、一阶导数[4]、二阶导数[6]、分数阶微分变换[6]等，有效减少甚至消除高光谱的冗余信息，选择合适的光谱变换方法能提升作物叶绿素反演精度。同样，植被指数法作为高光谱反演农作物生理生化指标的重要方法，也已被广泛应用于反演作物的生物化学参量[7-10]。随着机器学习算法快速发展，多元逐步回归(Multiple Linear Regression, MLR)[11]、偏最小二乘回归(Partial Least Squares Regression, PLSR)[12]、随机森林回归(Random Forests Regression, RFR)[1]、支持向量机回归(Support Vector Machine Regression, SVR)[13]等方法在估算叶片叶绿素含量研究中均取得了较好的效果。茹克亚等[6]利用对数变换和分数阶1.2阶微分计算的植被指数(NDVI、DVI、RVI、MSR_{705}、$MSR_{670,800}$、CI)建立的PLSR模型的估算精度最优。Liu等[5]利用连续小波变换对马铃薯叶绿素含量进行分析，结果表明连续小波变换方法能够有效的提取出对马铃薯叶绿素含量较为敏感且稳健的小波特征，增强了光谱与叶绿素含量之间的相关性。张思楠等[7]通过植被指数对多枝柽柳叶片叶绿素含量估算研究，选出了结构简单且表现较好的光谱指数，以上研究表明，植被指数法在反演作物叶片生理生化指标方面，能够取得较好的效果。而郭超凡等[14]提出的多植被指数协同估算模型精度相对于单植被指数模型精度明显提高。Qi等[15]对花生叶片叶绿素含量的高光谱反演结果表明，NDSI、RSI、DSI、SAVI是最佳的光谱指数。众多学者利用不同机器学习算法，不同光谱预处理方法，以及植被指数法对作物属性的定量监测进行了大量的研究并取得了显著的成果。因此，选择合适的光谱预处理方法、植被指数法与机器学习方法，对于利用光谱

收稿日期：2023-05-11； **修订日期：**2023-07-20
基金项目：新疆自然科学计划(自然科学基金)联合基金项目(2021D01C055)
作者简介：阿热孜古力·肉孜(1997-)，女，硕士研究生，主要从事干旱区资源环境及农业遥感应用方面的研究. E-mail: 173762855@qq.com
通讯作者：买买提·沙吾提. E-mail: korxat@xju.edu.cn

反演作物叶片叶绿素含量尤为重要。

综上所述,光谱数据变换方法和植被指数法广泛应用于作物参量估测研究中。但多数研究均是基于单一的光谱数据变换或单一植被指数构建模型,而多种光谱数据变换和多植被指数组合估测棉花叶绿素含量的研究鲜有报道。因此,本文结合前人研究选择46种在农作物参量研究方面应用较广的植被指数,基于传统数学变换,分数阶变换,小波变换,结合相关系数法与SVR、RFR和鲸鱼优化算法(Whale Optimization Algorithm, WOA)优化RFR建立多种植被指数回归模型,旨在解决以下问题:(1)光谱数据变换方法是否能提高植被指数与棉花叶片叶绿素含量的相关性。(2)探究最佳的植被指数组合和最佳估测模型。(3) WOA-RFR估测模型较RFR模型的优化效果。

1 材料与方法

1.1 研究区概况

研究区位于新疆阿克苏地区库车市阿克吾斯塘乡(82°58′~83°30′E,41°16′~41°31′N),地处天山南麓,塔里木盆地中北部,是典型的扇形平原绿洲。地形北高南低,自西北向东南倾斜,最高海拔高程为4550 m,最低海拔高程922 m。地处暖温带,气候干燥,降水稀少,夏季炎热,冬季干冷,年温差和日温差均较大,属暖温带大陆性干旱气候,年均温为10.5~11.4 ℃,年均降水量范围为43.1~51.6 mm[16],属于干旱与极端干旱区。研究区及样点分布情况见图1。

1.2 棉花样本采集和叶绿素含量的测定

棉花样本数据的采集于2021年5月26日至6月3日,在阿克苏地区库车市阿克吾斯塘乡的农用棉田进行,棉田总面积为0.5 hm^2,棉花品种是新陆中35号,棉花处于苗期阶段,共采集125个样本,采样点分布如图1所示,按照行距为70 cm,株距为10 cm的地膜覆盖垄种进行种植,棉田的施肥,栽培管理,按照当地种植要求正常进行。

选用叶绿素计(SPAD-502Plus)在获取高光谱数据的同时测定苗期棉花叶片SPAD值,叶绿素计(SPAD-502Plus)测得的读数与叶绿素含量有密切联系,可以将SPAD值转换为叶绿素含量进行估测[17],因此,本文将SPAD值等同于叶绿素含量进行研

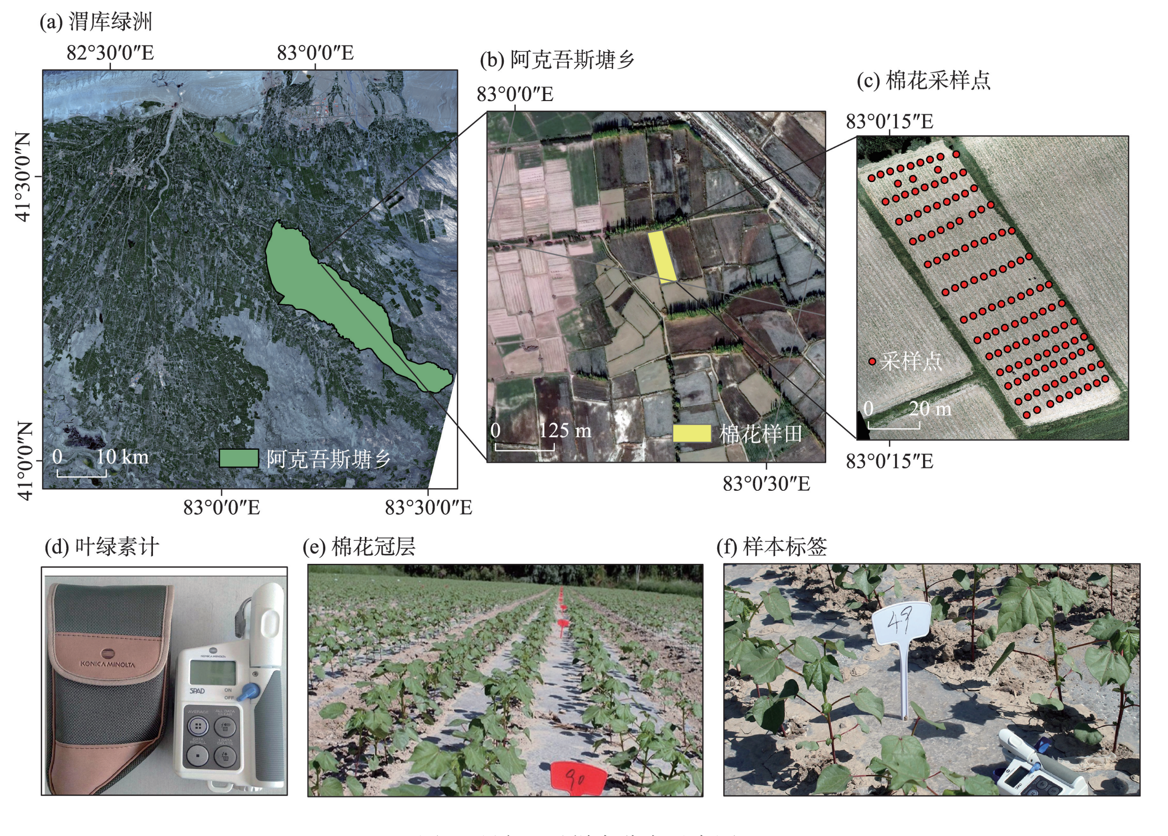


图1 研究区及样点分布示意图

Fig. 1 Schematic diagram of the distribution of the study area and sample points

究。测量每个样点上、中、下3层共6片叶子，测量每片叶子的不同部位6次，取其平均值作为最终值。

1.3 光谱数据采集与处理

棉花冠层光谱反射率使用ASD Field Spec Hand Held便携式光谱仪测定，该仪器的探测波长范围为325~1075 nm，光谱分辨率为3 nm。光谱测量时间为北京时间11:30—15:30，在无风无云晴朗的天气下进行。测量前先进行白板校准，测量时棉布冠层与光谱仪探头的距离为25 cm，并应垂直向下，光谱仪的视场角为25°，光谱扫描时间至少为8 s，每个采样点重复测量6次，每测量3次后进行一次白板校准。最后将在ViewSpec Pro软件中计算得出的每个采样点的6条曲线的平均值作为各采样点的光谱反射值。

本文对原波段光谱反射率（Original Reflectance，OR）进行对数变换（Logarithmic Transformation Spectrum, Log*R*）、平方根变换（Square Root Transformation, $\sqrt{R}$）、包络线去除（Continuum-removal Transformation, *CR*）、分数阶微分变换（Fractional-order Difference, FD）、连续小波变换（Continuous Wavelet Transform, CWT）等，其中传统的对数变换、平方根变换能够使可见光区域的光谱差异增强，使光谱的吸收、反射特征更加突出；包络线去除处理可压抑背景光谱，并扩大弱吸收特征信息[1]；分数阶微分变换有增多敏感波段以及提高相关系数的作用[6]；连续小波变换能够挖掘被隐藏有效光谱，提升光谱信息，增强光谱特征[5]。

1.4 植被指数的提取

本文结合前人的研究[18-25]，选取广泛应用于农作物叶绿素含量研究的46种植被指数（VI），作为叶绿素含量反演模型的参数，如表1所示。

1.5 模型构建和评价

本研究利用多元回归的支持向量机回归（SVR）模型和随机森林回归（RFR）模型分别进行建模[1-4]，比较其建模效果，并且采用鲸鱼算法（Whale Optimization Algorithm, WOA）优化RFR模型，检验模型优化效果，最终筛选出建模效果最好的模型。将125个样本分为两组，一组为建模集样本（*n*=82）用于模型的建立，一组为验证集样本（*n*=43）用于模型的验证。

模型的精度主要由决定系数（Determination Coefficients，R^2）、均方根误差（Root Mean Squared Error, RMSE）和平均相对误差（Mean Relative Error，MRE）三个指标来评价[24]，RMSE和MRE越小，说明模型的误差小，R^2越接近于1，说明模型的拟合效果越好。

鲸鱼优化算法是2016年提出来的一种新型智能优化算法，其源自于对自然界中座头鲸群体狩猎行为的模拟，通过鲸鱼群体搜索、包围、追捕、攻击猎物等过程实现优化搜索的目的[26]。该算法的计算方程及原理如下：

WOA寻优路径更新方法为：

$$D = |CX^*(t) - X(t)| \tag{1}$$

$$X(t+1) = X^*(t) - A \cdot D \tag{2}$$

式中：X^* 表示最优解坐标；X 为当前解坐标，系数向量 A 和 C 分别为

$$A = 2a \cdot r - a \tag{3}$$

$$C = 2r \tag{4}$$

式中：r 为[0,1]的随机向量。

设一座头鲸坐标为 X_{rand}，搜索猎物数学公式如下：

$$D = |C \cdot X_{rand} - X| \tag{5}$$

$$X(t+1) = X_{rand} - A \cdot D \tag{6}$$

螺旋攻击方法如公式(7)所示：

$$X(t+1) = D \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \tag{7}$$

当座头鲸发现猎物后，进行包围捕食或螺旋攻击，具体由选择概率 p 决定。

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D & p < 0.5 \\ D \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) & p \geqslant 0.5 \end{cases} \tag{8}$$

通过不断更新 $X(t+1)$，按照公式(8)进行猎食运动，循环迭代直至获取最优个体。

2 结果与分析

2.1 棉花叶片叶绿素含量统计与光谱特征分析

建模集和验证集中棉花叶片叶绿素含量的平均值分别为56.74和54.98，标准差分别为2.49和2.73，变异系数分别为4.4%和5.0%；全部样本的平均值为56.13，标准差为2.70，变异系数为4.8%，三个数据集变异系数均小于10%，数据变异性小(图2)。

为分析棉花叶片叶绿素含量与光谱反射率的响应特征，计算全部样本叶绿素含量的较高值、较低值、平均值和中间值，研究了不同叶绿素含量水

表1 高光谱植被指数

Tab. 1 Previously published hyperspectral vegetation indices

植被指数	计算公式	文献	植被指数	计算公式	文献
GNDVI	$(R_{800}-R_{500})/(R_{800}+R_{500})$	[19]	NRI	$(R_{570}-R_{670})/(R_{570}+R_{670})$	[20]
PSRI	$(R_{678}-R_{500})/R_{750}$	[20]	NPCI	$(R_{680}-R_{430})/(R_{680}+R_{430})$	[20]
VARI	$(R_{555}-R_{680})/(R_{580}+R_{680}+R_{480})$	[21]	RVI	R_{800}/R_{680}	[21]
NDVI	$R_{800}-R_{680}/R_{800}+R_{680}$	[21]	DVI	$R_{800}-R_{680}$	[21]
VOG1	R_{780}/R_{740}	[21]	VOG2	$R_{734}-R_{747}/R_{715}-R_{726}$	[21]
G	R_{780}/R_{740}	[23]	Lichtenthaler1	R_{440}/R_{690}	[23]
Lichtenthaler2	$(R_{800}-R_{680})/(R_{800}+R_{680})$	[23]	SIPI	$(R_{800}-R_{450})/(R_{800}+R_{450})$	[23]
PSSRa	R_{800}/R_{680}	[23]	Vogelmann1	$(R_{734}-R_{747})/(R_{715}+R_{720})$	[23]
Vogelmann3	R_{740}/R_{720}	[23]	ND	$(R_{935}-R_{705})/(R_{935}+R_{705})$	[23]
CIred edge	R_{750}/R_{705-1}	[24]	CIgreen	R_{800}/R_{550-1}	[24]
RI-half	R_{747}/R_{708}	[24]	GRVI	$(R_{670}/R_{500})-1$	[24]
RDVI	$(R_{800}-R_{670})/\sqrt{R_{800}-R_{670}}$	[25]	Datt1	$(R_{850}-R_{710})/(R_{850}+R_{680})$	[25]
Datt2	R_{850}/R_{710}	[25]	Datt3	R_{754}/R_{704}	[25]
Carte1	R_{695}/R_{420}	[25]	Carte2	R_{695}/R_{760}	[25]
Carte3	R_{605}/R_{760}	[25]	Carte4	R_{710}/R_{760}	[25]
$SR_{[752,690]}$	R_{752}/R_{690}	[25]	$SR_{[675,700]}$	R_{675}/R_{700}	[25]
NVI	$(R_{777}-R_{747})/R_{673}$	[25]	MSR705	$(R_{750}-R_{445})/(R_{705}+R_{445})$	[25]
SAVI	$(1+0.5)\times(R_{800}-R_{670})/(R_{800}+R_{670}+0.5)$				[19]
GARI	$[R_{800}-[R_{550}-1.7\times(R_{470}-R_{670})]][R_{800}+[R_{550}-1.7\times(R_{470}+R_{670})]]$				[19]
OSAVI	$1.16\times[(R_{800}-R_{670})/(R_{800}+R_{670}+0.16)]$				[21]
R_o	红谷(640 ~ 680 nm)内反射率最小值				[22]
mND_{705}	$(R_{750}-R_{705})/(R_{750}+R_{705}-2\times R_{445})$				[24]
EVI	$2.5\times[(R_{800}-R_{670})/(R_{800}+6R_{670}-7.5\times R_{475}+1)$				[25]
MCARI	$[(R_{700}-R_{670})-0.2\times(R_{700}-R_{550})]\times(R_{700}/R_{670})$				[25]
TVI	$0.5\times[120\times(R_{750}-R_{550})-2.5\times(R_{670}-R_{550})]$				[25]
MTVI1	$1.2\times[1.2\times(R_{800}-R_{550})-2.5\times(R_{670}-R_{550})$				[25]
REP	$700+40\times[(R_{670}+R_{780})/2-R_{700}]/(R_{740}-R_{700})$				[25]
SPVI	$0.4\times3.7\times(R_{670}-R_{550})-1.2\times\lvert R_{530}-R_{670}\rvert$				[25]
SPVI2	$0.4\times3.7\times(R_{800}-R_{670})-1.2\times\lvert R_{550}-R_{670}\rvert$				[25]

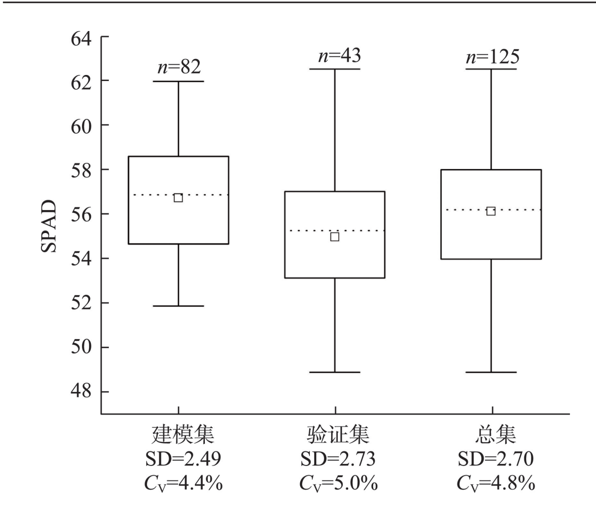
注：R_x代表x波段处的反射率。

平下的原始光谱反射率(图3)。不同叶绿素含量样本的光谱曲线的变化大体相似但也有一定的差别，这表明叶绿素含量的多少对光谱反射率有一定的影响，在可见光波段范围内550 nm附近形成了一个反射峰，690 nm附近形成了一个吸收谷，700~750 nm波段范围内，反射率急剧上升直至形成一个陡坡，950 nm处反射率逐渐下降，至975 nm又逐步上升。325~680 nm波段范围内叶绿素含量低反射率越高，这也证明了棉花叶片叶绿素含量对可见光波段的敏感性。

由图4可知，对数变换和包络线统去除变换的相关系数在全波段没有提高，相关性低于原始光谱，分数阶变换的相关系数(以FD-1.3为例)在700~800 nm相关性达到最高(r=0.41)，小波变换(以CWT-7为例)的相关性在600~700 nm呈负相关(r= −0.30)，在900~1000 nm呈正相关，相关系数最高达到(r=0.34)。对比不同数学光谱变换方法的Pearson相关系数，分数阶微分变换和小波变换是较好的预处理方法。

2.2 植被指数组合和筛选

本文运用46种植被指数，基于原始光谱、包络线去除、分数阶变换(以0.3为步长)、小波变换(分解尺度2^1、2^3、2^5、2^7、2^9)等处理的光谱数据上进行植被指数运算，选出若干个通过0.01相关性显著水平



注:SPAD为相对叶绿素含量;SD为标准差;C_v为变异系数。

图2　棉花叶片叶绿素含量统计

Fig. 2　Statistics on chlorophyll content in cotton leaves

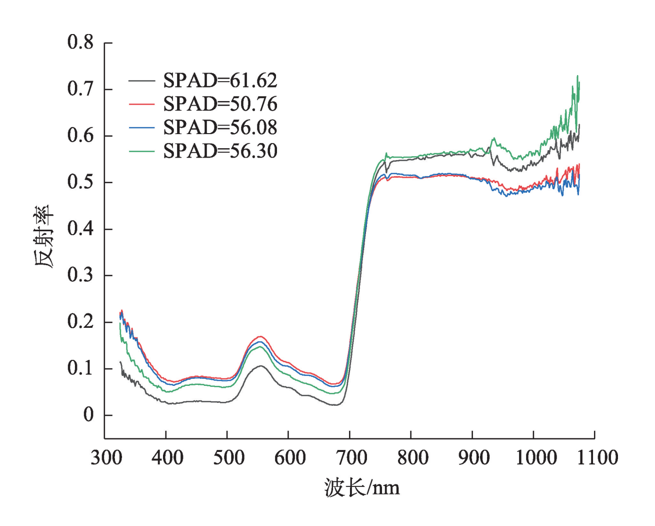
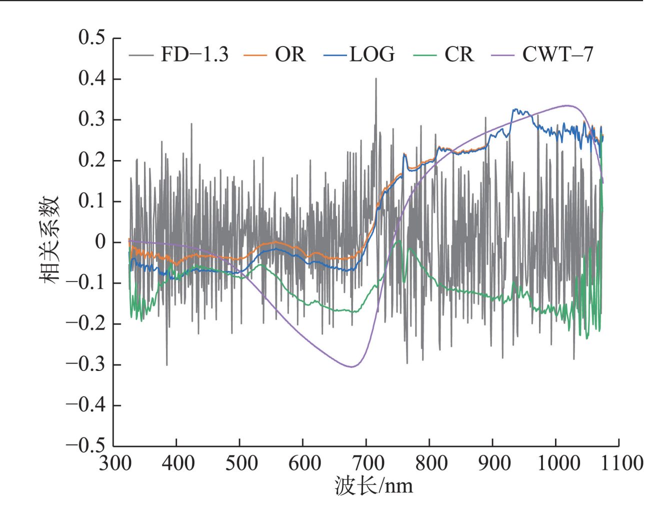


图3　不同叶绿素含量下光谱反射率变化

Fig. 3　Change in spectral reflectance under different chlorophyll content

的植被指数,作为一个组合植被指数(图5)。原始光谱数据上通过0.01相关性水平的植被指数只有6个,而传统数学变换方法中的对数变换,分数阶变换中的0.3、0.6、0.9阶,小波变换中的第3、第7尺度中通过0.01相关性显著水平的植被指数数量超过10个,说明这些数学变换形式提高了更多植被指数与棉花叶片叶绿素含量的相关性,对这些变换上的植被指数按照与叶绿素含量的相关性系数大小排序,取相关系数大的前10个植被指数作为一个组合植被指数,并且选择这些植被指数组合进行下一步分析。在平方根变换、包络线变换、分数阶的1.2阶、小波变换的第1、第5和第9尺度变换形式上通过0.01相关性显著水平的植被指数个数较少,分数



注:FD−0.3为分数阶1.3阶变换;OR为原始光谱反射率;LOG为对数光谱变换;*CR*为包络线去除;CWT−7为小波变换第7个尺度,下同。

图4　不同数学变换反射率与叶绿素含量的相关性分析

Fig. 4　Correlation analysis of reflectance and chlorophyll content in different mathematical transformations

阶1.5阶及之后的阶次上通过0.01相关性显著水平的植被指数个数均少于1,不做进一步分析。

2.3　多植被指数组合相关性分析

不同光谱变换下植被指数与棉花叶绿素含量相关性统计表,如表2所示,每一种光谱变换下的植被指数与棉花叶片叶绿素含量之间进行相关性分析,原始光谱下植被指数的相关性均低于光谱变换后植被指数的相关性,表明对光谱数据进行不同数学变换,能够提高植被指数与叶绿素含量的相关性,分数阶0.6阶次下植被指数的相关性为−0.33~0.34,小波变换的第7尺度下则为−0.31~0.34,证明分数阶变换和小波变换更能提高叶绿素含量和植被指数相关性。

2.4　建模结果与分析

利用多种光谱预处理方法和不同植被指数组合,分别建立SVR、RFR以及WOA-RFR模型估算棉花叶片叶绿素含量,研究不同植被指数组合方法对模型精度的影响,其结果如表3所示。从表中可知,在不同数学变换上RFR、WOA-RFR模型效果较好,且WOA-RFR优于RFR模型,证明了鲸鱼优化算法的可行性,而SVR模型的反演效果相对较差,SVR模型在建模集和验证集的模型精度差异较大,这说明该模型受样本个数影响较大,模型稳定性差。基于对数变换多植被指数组合的RFR和WOA-RFR模型的训练集R^2分别为0.884和0.912,验证集R^2分别

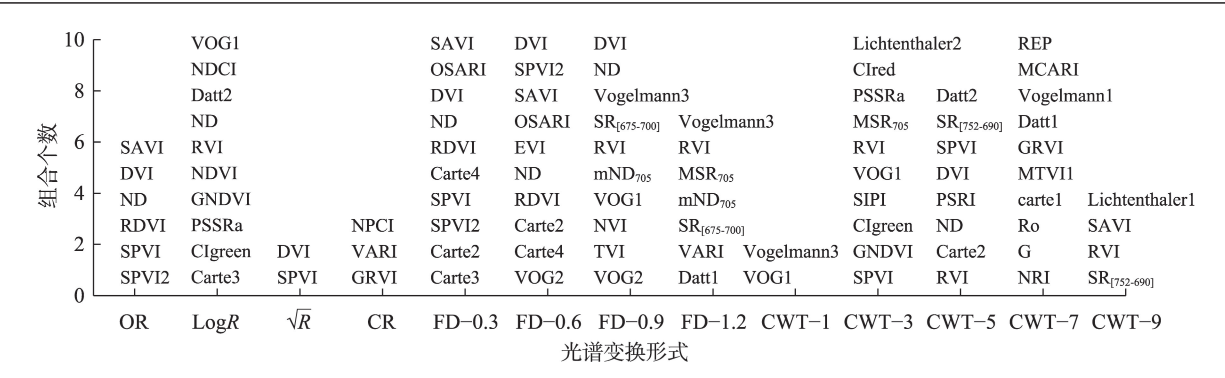


图5 不同数学变换下的植被指数组合

Fig. 5 Vegetation index combination under different mathematical transformations

表2 棉花叶片叶绿素含量与不同光谱变换下多植被指数的相关性分析

Tab. 2 Correlation analysis of chlorophyll content and vegetation index under different spectral transformations of cotton leaves

光谱变换	植被指数	相关性范围
OR	SAVI,DVI,ND,RDVI,SPVI,SPVI2	0.24~0.26
Log*R*	VOG1,NDCI,Datt2,ND,RVI,NDVI,GNDVI,PSSRa,CIgreen,carte3	−0.27~−0.37
FD−0.3	SAVI,OSARI,DVI,ND,RDVI,Carte3,Carte4,Carte2,SPVI,SPVI2	−0.28~0.29
FD−0.6	DVI,SPVI2,SAVI,OSARI,EVI,ND,RDVI,Carte2,Carte4,VOG2	−0.33~0.34
FD−0.9	Vogelmann3,RVI,DVI,SR$_{[675-700]}$,Mndvi$_{705}$,ND,VOG1,NVI,TVI,VOG2	−0.23~0.30
CWT−3	CIred,PSSRa,MSR$_{705}$,RVI,Lichtenthaler2,VOG1,SIPI,CIgreen,GNDVI,SPVI	−0.27~0.28
CWT−7	REP,MCARI,Vogelmann1,Datt1,GRVI,MTVI1,carte1,Ro,G,NRI	−0.31~0.34

注:OR为原始光谱反射率;Log*R*为对数光谱变换;FD−0.3为分数阶0.3阶变换;CWT−3为小波变换第3个尺度,下同。

为0.858和0.894。分数阶变换中,0.9阶的RFR模型的训练集R^2为0.916,验证集R^2为0.949,均高于其他阶次RFR模型的R^2,0.9阶的WOA-RFR模型的训练集R^2为0.920,验证集R^2为0.955,比原始光谱植被指数组合WOA-RFR模型的R^2提高了4.5%。在小波变换尺度上,第7尺度上的WOA-RFR模型的训练集R^2最高(R^2=0.934),综合分析,上述模型中基于分数阶0.9阶变换上的多植被指数组合建模模型WOA-RFR的R^2高,RMSE和MRE低,建模效果最佳,可以较为准确的估测棉花叶片叶绿素含量。

从预测值和实测值的拟合图可以看出(图6),不同数学变换下验证集的模型预测效果比建模集的预测效果好,基于分数阶0.9阶次的组合植被指数建立的模型预测效果好,除个别数据点其他都较紧密分布在拟合线周围,拟合效果较好。

3 讨 论

叶绿素含量一直是评价农作物长势、产量和质量的重要生化参数,利用光谱技术对作物的叶绿素含量进行定量反演是精准农业研究的热点。鉴于此,本文以南疆棉花为研究对象,基于不同的光谱数据变换方法计算植被指数,研究结果表明,数据变换方法能够提高植被指数与叶绿素含量的相关性,这与前人的研究结果一致,童新等[27]基于不同阶微分高光谱植被指数的牧区草场地上生物量估算研究中表明对原始高光谱反射率进行微分处理,有助于光谱数据与地上生物量的相关性;李莉婕等[28]通过高光谱定量反演火龙果茎枝叶绿素含量的研究得出数学变换和连续小波变换能明显提升光谱对火龙果茎枝叶绿素含量的估测能力的结论;竞霞等[29]对小麦条锈病遥感监测模型构建的研究中通过对原始光谱进行分数阶微分处理,提高了光谱反射率与小麦条锈病严重度的相关性。另外,本研究结果表明,利用多种植被指数组合构建棉花叶片叶绿素含量的模型精度较其他研究高,在所有光谱数据变换上RFR和WOA-RFR模型精度都高于0.85,而在分数阶微分的0.9阶DVI、ND、Vogel-

表3 模型建模结果比较

Tab. 3 Comparison of model modeling results

光谱变换	模型算法	建模集			验证集		
		R^2	RMSE	MRE	R^2	RMSE	MRE
OR	WOA-RFR	0.889	0.014	0.015	0.918	0.050	0.015
	RFR	0.860	1.307	0.019	0.862	1.240	0.018
	SVR	0.756	1.241	0.013	0.908	0.854	0.006
Log*R*	WOA-RFR	0.912	1.067	0.016	0.894	1.081	0.015
	RFR	0.884	1.367	0.020	0.858	1.280	0.020
	SVR	0.658	1.460	0.015	0.943	0.651	0.006
FD-0.3	WOA-RFR	0.917	1.019	0.014	0.878	1.060	0.016
	RFR	0.872	1.357	0.020	0.850	1.269	0.019
	SVR	0.609	1.569	0.017	0.866	1.035	0.010
FD-0.6	WOA-RFR	0.890	1.085	0.015	0.925	1.008	0.014
	RFR	0.872	1.379	0.019	0.846	1.388	0.021
	SVR	0.451	1.836	0.020	0.685	1.594	0.018
FD-0.9	WOA-RFR	0.920	0.987	0.013	0.955	0.986	0.014
	RFR	0.916	1.250	0.018	0.949	1.207	0.017
	SVR	0.754	1.263	0.013	0.882	0.973	0.010
CWT-3	WOA-RFR	0.895	1.091	0.015	0.922	1.086	0.014
	RFR	0.892	1.388	0.020	0.900	1.443	0.020
	SVR	0.779	1.172	0.010	0.906	0.835	0.007
CWT-7	WOA-RFR	0.934	0.946	0.013	0.911	1.082	0.015
	RFR	0.913	1.286	0.019	0.841	1.281	0.020
	SVR	0.522	1.724	0.019	0.791	1.254	0.009

注:R^2为决定系数;RMSE为均方根误差;MRE为平均相对误差;WOA为鲸鱼优化算法;RFR为随机森林回归模型;SVR为支持向量机回归模型。

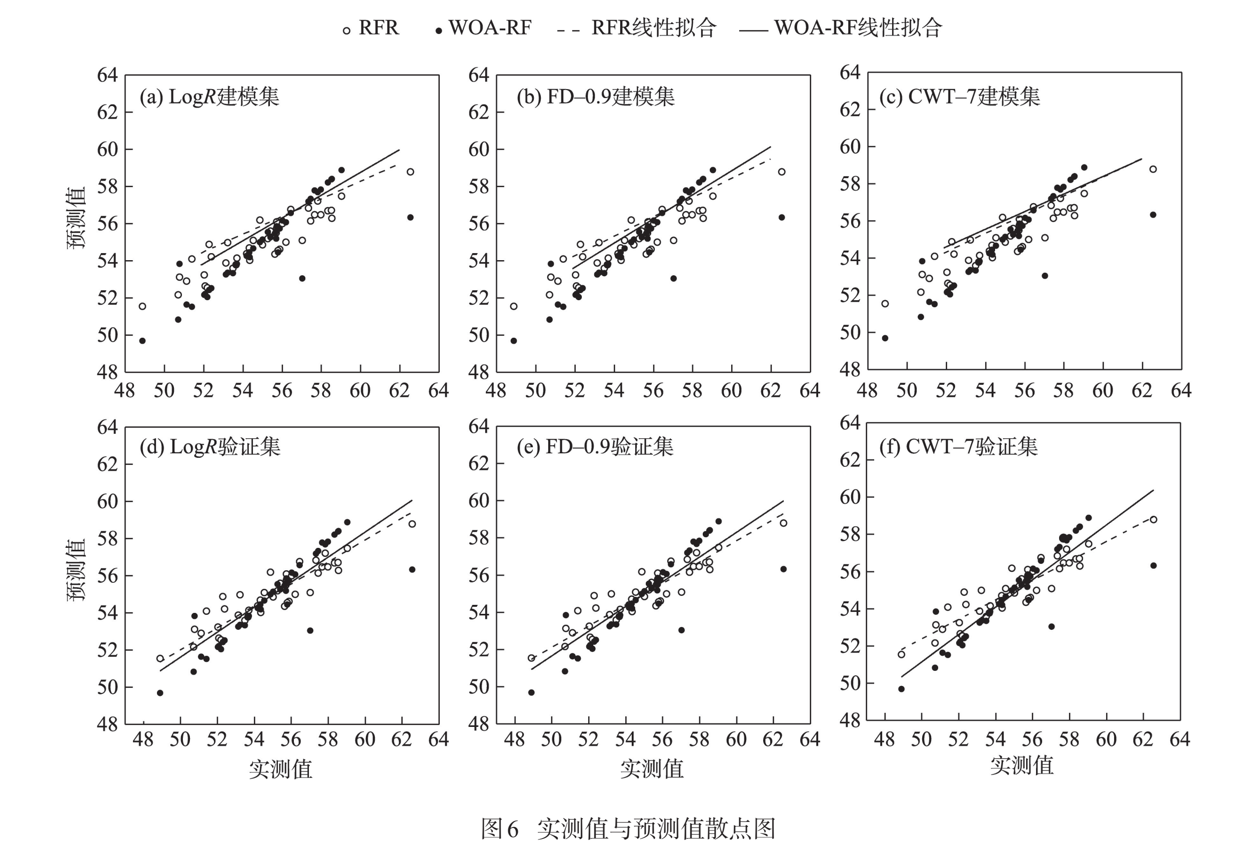


图6 实测值与预测值散点图

Fig. 6 Scatter plots of measured and predicted value

mann3、$SR_{[675-700]}$、RVI、mND_{705}、VOG1、NVI、TVI、VOG2的多植被指数组合的WOA-RF反演模型精度达到0.920。Wu等[30]的研究证明了由特定波段的反射率组成的高光谱植被指数反演叶绿素含量的可行性；许童羽等[18]基于多植被指数组合的BAS-ELM粳稻冠层氮含量反演方法的研究中9组不同输入量建立的BAS-ELM反演建模效果均较好，预测效果最好的是以8种植被指数的组合为输入量建立的模型，训练集与验证集的R^2分别为0.756和0.753，优于单一结构的植被指数组合。但是，本文只对棉花苗期的叶片叶绿素含量进行了反演研究，但叶绿素含量是会受到不同生育期的影响，其会随着作物长势的变化而变化，因此，在今后研究中加强该模型在不同生育期的棉花生理生化参数应用。对于不同的农作物及其不同的生育期内植被指数适用性不同，生育时期会影响植被指数的叶绿素含量估测能力，因此，在今后的研究中有必要对植被指数进行优化，进一步探究其在叶绿素含量反演研究中的潜力。

4 结 论

（1）对原始光谱进行变换都能提高叶绿素含量和植被指数的相关性，传统变换在对数变换时相关性最高，分数阶变换在0.9阶时相关性最高，小波变换在第7个尺度上相关性最高。在本研究中分数阶变换比小波变换更能提高叶绿素含量和植被指数的相关性。

（2）基于不同光谱数据变换的不同植被指数组合的反演模型效果较好，均高于0.7，基于分数阶0.9阶变换Vogelmann3、RVI、DVI、$SR_{[675-700]}$、$Mndvi_{705}$、ND、VOG1、NVI、TVI、VOG2多植被指数组合的WOA-RFR模型反演效果最佳，其建模集和验证集模型R^2分别为0.920和0.955，RMSE分别为0.987和0.986，MRE分别为0.013和0.014。

（3）在所有的反演模型中，RFR模型优于SVR模型，而WOA算法优化效果较明显，对每一种光谱变换的植被指数组合的RFR模型都有一定的优化，R^2至少提高2%，最大提高6.6%，RMSE至少降低30%，鲸鱼优化算法对随机森林回归模型有较高的普适性。

参考文献（References）：

[1] 依尔夏提·阿不来提, 买买提·沙吾提, 白灯莎·买买提艾力, 等. 基于随机森林法的棉花叶片叶绿素含量估算[J]. 作物学报, 2019, 45(1): 81–90. [Ershat Ablet, Mamat Sawut, Baidengsha Maimaitiaili, et al. Estimation of leaf chlorophyll content in cotton based on the random forest approach[J]. Acta Agronomica Sinica, 2019, 45(1): 81–90.]

[2] 王伟东, 常庆瑞, 王玉娜. 基于UHD185成像光谱仪的冬小麦叶绿素监测[J]. 干旱区研究, 2020, 37(5): 1362–1369. [Wang Weidong, Chang Qingrui, Wang Yuna. SPAD monitoring of winter wheat based on a UHD185 imaging spectrometer[J]. Arid Zone Research, 2020, 37(5): 1362–1369.]

[3] Xiao Q L, Tang W T, Zhang C, et al. Spectral preprocessing combined with deep transfer learning to evaluate chlorophyll content in cotton leaves[J]. Plant Phenomics, 2022, 2022: 9813841.

[4] 安赛日丁·艾尔肯, 买买提·沙吾提, 李金朝. 基于冠层光谱特征的长绒棉叶绿素含量估算[J]. 激光与光电子学进展, 2022, 59(5): 1–11. [Arkin Ansardin, Sawut Mamat, Li Jinzhao. Estimation of chlorophyll content of long-staple cotton based on canopy spectrum characteristics[J]. Laser & Optoelectronics Progress, 2022, 59(5): 1–11.]

[5] Liu N, Xing Z P, Zhao R M, et al. Analysis of chlorophyll concentration in potato crop by coupling continuous wavelet transform and spectral variable optimization[J]. Remote Sensing, 2020, 12(17): 2826.

[6] 茹克亚·萨吾提, 阿不都艾尼·阿不里, 尼加提·卡斯木, 等. 基于分数阶微分的春小麦叶绿素含量高光谱估算[J]. 麦类作物学报, 2019, 39(6): 738–746. [Rukeya Sawuti, Abuaini Abula, Nijati Kasimu, et al. Spectral estimation of chlorophyll content in spring wheat leaves based on fractional differential[J]. Journal of Triticeae Crops, 2019, 39(6): 738–746.]

[7] 张思楠, 王权, 靳佳, 等. 应用光谱指数法估算多枝柽柳同化枝叶绿素含量[J]. 干旱区研究, 2016, 33(5): 1088–1097. [Zhang Sinan, Wang Quan, Jin Jia, et al. Application of hyperspectral indices for estimating leaf chlorophyll content of assimilating shoots of Tamarix Ramosissima[J]. Arid Zone Research, 2016, 33(5): 1088–1097.]

[8] 李长春, 施锦锦, 马春艳, 等. 基于小波变换和分数阶微分的冬小麦叶绿素含量估算[J]. 农业机械学报, 2021, 52(8): 172–182. [Li Changchun, Shi Jinjin, Ma Chunyan, et al. Estimation of chlorophyll content in winter wheat based on wavelet transform and fractional differential[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(8): 172–182.]

[9] Verma B, Prasad R, Srivastava P, et al. Investigation of optimal vegetation indices for retrieval of leaf chlorophyll and leaf area index using enhanced learning algorithms[J]. Computers and Electronics in Agriculture, 2022, 192: 106581.

[10] Wu C Y, Zheng N, Tang Q, et al. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation[J]. Agricultural and Forest Meteorology, 2008, 148(8–9): 1230–1241.

[11] 张卓然. 棉花高光谱特征及其农学参数遥感反演研究[D]. 杨凌: 西北农林科技大学, 2018. [Zhang Zhuoran. Research on Hy-

perspectral Characteristic of Cotton and Remote Sensing Inversion about Cotton Agronomic Parameters[D]. Yangling: Northwest A&F University, 2018.]

[12] Liu N F, Townsend P A, Naber M R, et al. Hyperspectral imagery to monitor crop nutrient status within and across growing seasons [J]. Remote Sensing of Environment, 2021, 255: 112303.

[13] Sun Q, Jiao Q J, Chen X D, et al. Machine learning algorithms for the retrieval of canopy chlorophyll content and leaf area index of crops using the PROSAIL-D model with the adjusted average leaf angle[J]. Remote Sensing, 2023, 15(9): 2264.

[14] 郭超凡, 陈雯璟, 牛明艳, 等. 基于多植被指数模型的草地地上生物量协同估算[J]. 干旱地区农业研究, 2022, 40(4): 206-213. [Guo Chaofan, Chen Wenjing, Niu Mingyan, et al. Collaborative estimation of aboveground biomass in grassland based on multiple vegetation index models[J]. Agricultural Research in the Arid Areas, 2022, 40(4): 206-213.]

[15] Qi H X, Zhu B Y, Kong L X, et al. Hyperspectral inversion model of chlorophyll content in peanut leaves[J]. Applied Sciences, 2020, 10(7): 2259.

[16] 玉苏甫·买买提, 吐尔逊·艾山, 买合皮热提·吾拉木. 新疆渭-库绿洲棉花种植面积遥感监测研究[J]. 农业现代化研究, 2014, 35(2): 240-243. [Yusup Mamat, Tursun Hasan, Magpirat Gulam. Remote sensing of cotton plantation areas monitoring in delta oasis of Ugan-Kucha River, Xinjiang[J]. Research of Agricultural Modernization, 2014, 35(2): 240-243.]

[17] 刘楠, 杨海波, 高飞, 等. 基于查找表法和优化光谱指数的马铃薯叶绿素反演[J]. 中国马铃薯, 2022, 36(6): 495-507. [Liu Nan, Yang Haibo, Gao Fei, et al. Potato chlorophyll inversion based on LUT and optimized spectral index[J]. Chinese Potato Journal, 2022, 36(6): 495-507.]

[18] 许童羽, 邢思敏, 于丰华, 等. 基于多植被指数组合的BAS-ELM粳稻冠层氮含量反演方法[J]. 沈阳农业大学学报, 2021, 52(5): 577-585. [Xu Tongyu, Xing Simin, Yu Fenghua, et al. A BAS-ELM inversion method of Japonica rice canopy nitrogen content based on A combination of multiple vegetation indices[J]. Journal of Shenyang Agricultural University, 2021, 52(5): 577-585.]

[19] 陈圣波, 陈彦冰, 任枫荻, 等. 基于光谱指数的玉米叶绿素含量估算[J]. 信阳师范学院学报(自然科学版), 2021, 34(2): 225-229. [Chen Shengbo, Chen Yanbing, Ren Fengdi, et al. Estimation of maize chlorophyll content based on spectral index[J]. Journal of Xinyang Normal University(Natural Science Edition), 2021, 34(2): 225-229.]

[20] 傅坤亚, 张文宇, 曹宏鑫, 等. 基于光谱的作物病虫害监测研究进展[J]. 中国农业科技导报, 2014, 16(5): 90-98. [Fu Kunya, Zhang Wenyu, Cao Hongxin, et al. Research progress on crop diseases and insect pests monitoring based on spectrum[J]. Journal of Agricultural Science and Technology, 2014, 16(5): 90-98.]

[21] 王芳东, 严志雁, 赵小敏, 等. 油茶叶片叶绿素含量高光谱估测的偏最小二乘模型参数选择[J]. 江西农业大学学报, 2022, 44(1): 86-96. [Wang Fangdong, Yan Zhiyan, Zhao Xiaomin, et al. Parameter selection of partial least squares model for hyperspectral estimation of chlorophyll content in camellia oleifera leaves[J]. Acta Agriculturae Universitatis Jiangxiensis, 2022, 44(1): 86-96.]

[22] 王鑫梅, 张劲松, 孟平, 等. 基于无人机遥感影像的核桃冠层氮素含量估算[J]. 农业机械学报, 2021, 52(2): 178-187. [Wang Xinmei, Zhang Jinsong, Meng Ping, et al. Estimation of nitrogen content in walnut canopy based on UAV remote sensing image[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(2): 178-187.]

[23] 孙林, 程丽娟. 植被叶片生化组分的光谱响应特征分析[J]. 光谱学与光谱分析, 2010, 30(11): 3031-3035. [Sun Lin, Cheng Lijuan. Analysis of spectral response of vegetation leaf biochemical components[J]. Spectroscopy and Spectral Analysis, 2010, 30(11): 3031-3035.]

[24] 洪帅, 张泽, 张立福, 等. 滴灌棉花不同生育时期冠层叶片叶绿素含量的高光谱估测模型[J]. 棉花学报, 2019, 31(2): 138-146. [Hong Shuai, Zhang Ze, Zhang Lifu, et al. Hyperspectral estimation model of chlorophyll content in cotton canopy leaves under drip irrigation at different growth stages[J]. Cotton Science, 2019, 31(2): 138-146.]

[25] Liang L, Di L P, Zhang L P, et al. Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method[J]. Remote Sensing of Environment, 2015, 165: 123-134.

[26] Mirjalili S, Lewis A. The whale optimization algorithm[J]. Advances in Engineering Software, 2016, 95: 51-67.

[27] 童新, 杨震雷, 张亦然, 等. 基于不同阶微分高光谱植被指数的牧区草场地上生物量估算[J]. 草地学报, 2022, 30(9): 2438-2448. [Tong Xin, Yang Zhenglei, Zhang Yiran, et al. Estimation of pasture aboveground biomass using different orders of differential hyperspectral vegetation indices[J]. Acta Agrestia Sinica, 2022, 30(9): 2438-2448.]

[28] 李莉婕, 岳延滨, 王延仓, 等. 高光谱定量反演火龙果茎枝叶绿素含量的研究[J]. 光谱学与光谱分析, 2021, 41(11): 3538-3544. [Li Lijie, Yue Yanbin, Wang Yancang, et al. The quantitative study on chlorophyll content of Hylocereus Polyrhizus based on hyperspectral analysis[J]. Spectroscopy and Spectral Analysis, 2021, 41(11): 3538-3544.]

[29] 竞霞, 张腾, 邹琴, 等. 基于分数阶微分光谱指数的小麦条锈病遥感监测模型构建[J]. 农业工程学报, 2021, 37(17): 142-151. [Jing Xia, Zhang Teng, Zou Qin, et al. Construction of remote sensing monitoring model of wheat stripe rust based on fractional-order differential spectral index[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(17): 142-151.]

[30] Wu C Y, Niu Z, Tang Q, et al. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation[J]. Agricultural and Forest Meteorology, 2008, 148(8): 1230-1241.

Estimation of cotton leaf chlorophyll content based on combinations of multi-vegetation indices

Areziguli ROZI[1,2,3], Mamat SAWUT[1,2,3], HE Xugang[1,2,3], YE Xiaowen[1,2,3]

(1. College of Geography and Remote Sensing Sciences, Xinjiang University, Urumqi 830017, Xinjiang, China; 2. Xinjiang Key Laboratory of Oasis Ecology, Urumqi 830017, Xinjiang, China; 3. Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, Urumqi 830017, Xinjiang, China)

Abstract: Chlorophyll content is a crucial indicator for characterizing vegetation growth. In this study, we utilized high-spectral technology to rapidly monitor the chlorophyll contents of cotton leaves. We collected 125 cotton leaf seedling samples from Xinjiang and measured their chlorophyll content and spectral data. To achieve this, we employed various spectral preprocessing techniques and used a combination of vegetation indices. Subsequently, we constructed a whale optimization algorithm/random forest regression (WOA-RFR) quantitative inversion model for cotton leaf chlorophyll content. Finally, we conducted a comparative analysis, contrasting the results of the WOA-RFR model with those obtained from the support vector regression (SVR) and RFR models. The results indicated that the spectral transformation methods (logarithm transformation, fractional order differentiation, and wavelet transformation) effectively improved the correlation between the vegetation indices and the chlorophyll content. We also found that the best inversion performance was achieved with the WOA-RFR model using a fractional order differentiation with a transformation order of 0.9 and the Vogelmann3, RVI, DVI, $SR_{[675-700]}$, $Mndvi_{705}$, ND, VOG1, NVI, TVI, VOG2 combined vegetation indices. The model exhibited R^2 values of 0.920 and 0.955 for the training set and validation set, respectively. The corresponding RMSE values were 0.987 and 0.986, while the MRE values were 0.013 and 0.014. Compared to the RFR and SVR models, the WOA-RFR model demonstrated higher predictive accuracy, and the optimization effect of the WOA algorithm was evident. As a result, this study provides valuable decision-making support for accurately quantifying cotton leaf chlorophyll content.

Keywords: combination of vegetation index; cotton; chlorophyll content; whale optimization algorithm